Sonar Transient False Alarm Reduction based on Automatic Detection and Characterization of Marine Mammal Sounds

Antonio Sánchez-García*, F. Javier Rodrigo*, José Luis Sancho-Gómez[+]

* *Sociedad Anónima de Electrónica Submarina (SAES), Carretera de la Algameca, s/n, 30205- Cartagena (Spain)*

a.sanchez@electronica-submarina.com, f.rodrigo@electronica-submarina.com

[+] *Universidad Politécnica de Cartagena. Departamento de Tecnologías de la Información y las Comunicaciones*

Campus Muralla del Mar s/n, 30202-Cartagena (Spain).

Josel.sancho@upct.es

Abstract

Methods for automatic detection of marine mammal signals can be incorporated to on-board sonars as a mean to separate these transient signals from those corresponding to contacts in order to reduce their false alarm rate. This detection configures likewise as the primary mean to preserve marine mammals if measures are taken in the area where the detection took place, to avoid high-level man-made noise that could potentially disturb or even damage these animals. On this basis, a method for the automatic detection and characterization of marine mammal signals is presented. In this method, whistle signals are all automatically detected from the spectrogram, isolated using region-growing segmentation, extracted and finally characterized by means of a fixed number of Radial Basis Function (RBF) coefficients. These coefficients can be later used to classify signals based on their characteristics. The performance of the method has been tested using synthetically generated and recorded at sea signals.

Keywords: *Detection, characterization, frequency contour, segmentation, RBF, marine mammal.*

1 INTRODUCTION

Short-duration transient signals are becoming an increasingly important mean for the sonar-based detection and identification of the threat as the source level of the broadband and narrowband long-duration acoustic signals emitted by vessels is continuously decreasing. Transient signals in the ocean have a wide variety of origins including Anthropogenic, biological and vessels. Short-duration anthropogenic noise is mainly centred in coastal areas and therefore, in the open sea, the remaining main sources of transient signals are biological and those produced by vessels.

The detection of transient signals on-board of vessels by means of sonar processing has two direct important consequences: On the one side, the capability of discriminating between signals emitted by biological sources from those emitted by targets, permits to decrease the false alarm percentage in transient sonar processing, and on the other hand, this identification also allows the vessel's crew to take measures in order to avoid the generation of high level emissions that could potentially disturb or damage the animals in its proximity. Nowadays, an increasing concern exists in relation with this potential effect on marine mammals, and different countries are establishing regulations in order to protect them [1].

Cetaceans together with sirenians (manatees and dugons) and pinnipeds (seals, sea lions and walrus) form the three orders of today living marine mammals. Cetaceans are the whales, dolphins and porpoises and comprise more than 80 species [2]. They are vocalizing animals that inhabit in all the seas and oceans of the world. Their vocal repertoire is composed by broadband and narrowband transient signals. Broadband signals are pulsed sounds generically referred to as clicks and can be emitted in trains containing few to hundreds of them or as burst pulse calls that are click trains with very short inter-click intervals which are not individually distinguishable by the human auditory system. Peak frequencies vary from tens of kilohertz to well over 100 kHz [3]. Narrowband signals are tonal sounds that usually correspond to frequency modulated whistles that often have harmonic components. They range in duration from several tenths of a second to several seconds. Their fundamental frequency ranges from 2 to 30 kHz [4]. A pod of bottlenose dolphins is shown in Figure 1.

This paper centres itself on frequency modulated whistles emitted by delphinids and presents a first approach to an automated method for detecting and characterizing these sounds. The characterization step provides as output, the weights associated to each of the radial basis functions used to approximate the extracted frequency contour corresponding to each whistle. These weights can be later on used to classify the whistles based on selected premises by using a neural network with a number of input neurons equal to the number of RBFs.

Automatic methods for detecting transient signals configure as an effective mean to reduce the operator workload. This reduction is significantly important for both, the sonar operator and the people in charge of specifically detecting the presence of marine mammals. The former, has to face with the increasing number of functionalities incorporated in modern sonars and the latter will benefit either from the no necessity of performing a continuous surveillance of the captured data in the case of real-time detection or of checking the whole range of recordings in the case of later analysis.

This paper first presents a background on region-based segmentation and Radial Basis Functions. Then, a description of the different steps making part of the processing chain is included, followed by an evaluation of the results obtained by first using synthetically generated signals and later using signals recorded at sea. Finally the conclusions of the study are presented.

Figure 1. Bottlenose dolphin pod

2 BACKGROUND

Historically, underwater sounds have been processed using predominantly signal-processing techniques. Nowadays, new disciplines are progressively being incorporated in order to deal with new and more stringent requirements. Among these new disciplines, image processing and neural network processing can be stood out. Image processing techniques are particularly well suited to work in combination with the spectrogram time-frequency representation due to the fact that the spectrogram output is a 2D matrix holding time-frequency cells. This matrix can be directly mapped to an image and then, apply the wide set of image processing functions already developed or new ones specifically designed, for the particular case to be addressed. On the other hand, neural network processing is an effective well-proven technique for the characterization and classification of data sets that can be separated in regions.

Next, a background on the image processing based segmentation technique for extracting individual objects and on the Radial Basis Functions is presented.

2.1 REGION-BASED SEGMENTATION

Image segmentation is a technique to subdivide an image into its constituent regions. It is currently applied to a wide number of disciplines as: medical imaging, face and fingerprint recognition, locating images in satellite images, automatic traffic controlling systems or artificial vision.

Image segmentation algorithms base generally on two basic properties of intensity images: discontinuity and similarity. In the former, the approach to partition an image is to search for abrupt changes in intensity, as these correspond to edges. In the latter, the approach is to find similar regions according to a set of predefined criteria. Some examples of methods included in this latter category are: thresholding, region growing and splitting and merging [5].

Segmentation resulting from region-growing methods perform generally better than those based on edge methods in noise images, where borders are extremely difficult to detect.

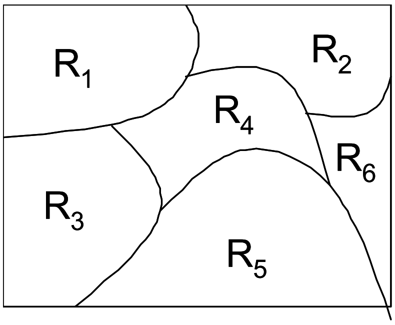

Figure 2. Division of a region in subregions

If we denote *R* to an entire image region, region-based segmentation can be viewed as a process that partitions *R* into subregions R_1, $R_2, \ldots, Rn$ (see Figure 2) based on the following conditions:

a) Completeness. Each pixel must be in a region.

$$\bigcup_{i=1}^{n} R_i = R$$

b) Connectedness. Each pixel in a region must be connected in some predefined sense.

R_i is a connected region, $i = 1,2,\ldots,n$.

c) Regions must be disjoint.

$R_i \cap R_j = \varnothing$ for all i and j, $i \neq j$

where $\varnothing$ is the null set.

d) Each region must be homogeneous.

$P(R_i) = TRUE$ for $i = 1,2,\ldots,n$

where $P(R_i)$ is a logical predicate, that is, a property that the pixel values of region R_i satisfy.

e) Each union of adjacent regions must be inhomogeneous.

$$P(R_i \cup R_j) = FALSE \quad \text{for all adjacent } R_i, R_j, i \neq j$$

The selected region-based segmentation method for the proposed method is region growing. It is a procedure that groups pixels or subregions into larger regions based on predefined criteria for growth. The basic approach consists on starting with a set of 'seed' points, and from these grow regions by appending to each seed neighbouring pixels that have predefined properties similar to the seed [5]. Growing process continues until a stopping criterion is satisfied.

Main considerations in region growing are the selection of a set of one or more starting points, the selection of the similarity criterion to add new pixels to the region, based on intensity levels or spatial properties, and the formulation of a stopping rule that determines when the growing process has been completed for each region (no more pixels satisfy the similarity criterion). This can take advantage of structural information, such as region shape or size, if it is available.

2.2 RADIAL BASIS FUNCTIONS (RBF)

A radial basis function is a real-valued function whose value depends on the distance from the origin or alternatively on the distance from some other point *m* called a centre.

Radial basis functions are typically used to build up approximation function of the form:

$$F(x) = \sum_{i=1}^{N} \omega_i \Phi(\|x - m_i\|) \tag{1}$$

where *F(x)* is the approximating function that is computed as the sum of N radial basis functions with a different centre m_i and weighed by a coefficient ω_i.

The approximation function in (1) admits a parallel structure as a neural network that is named as radial basis function network. This network is characterized by a fixed structure composed by three layers with entirely different roles [6]. The input layer is made up of source nodes that connect the network to its environment. The second layer, that constitutes the only hidden layer of the network, applies a nonlinear transformation from the input space to the output space, and the third layer is the output layer that is linear and supplies the response of the network to the activation pattern applied to the input layer.

The RBF network is a feed-forward net where all the nodes are completely connected. A schema of this network is presented in Figure 3.

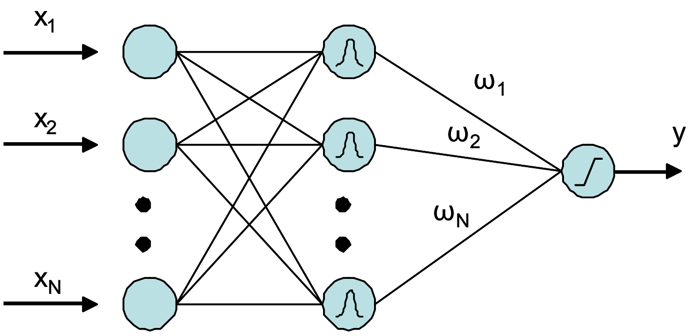

Figure 3. Structure of a radial basis function network composed by three fixed layers

The most commonly used radial basis function is the Gaussian kernel, defined as follows:

$$\Phi(\|x-m\|) = \exp\left(-\frac{\|x-m\|^2}{2\sigma^2}\right) \qquad (2)$$

where m is the centre of the RBF and $2\sigma^2$ its width.

Training an RBF network consists on determining its unknown parameters. This means determining (a) the number of radial basis functions that corresponds to the number of hidden units, (b) the centres and widths of each radial basis function, and (c) the weights of the output layer. Different approaches have been proposed to determine these parameters. The presented method considers a fixed number of radial basis functions with equally spaced centres and computes the weights of the output layer by means of linear least square, using a pseudo inverse matrix to prevent singularities.

3 PROPOSED METHOD

The proposed method covers the first two steps of a pattern recognition system: detection and feature extraction or characterization. Its main stages are presented in Figure 4.

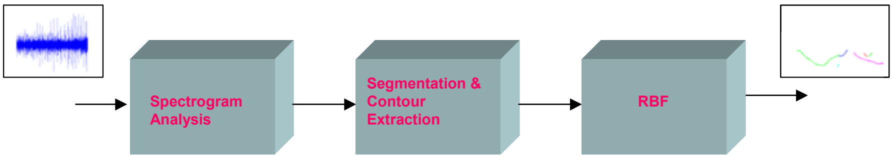

Figure 4. Main stages of the proposed method

The first stage performs the spectrogram analysis that converts time-domain input data into a time-frequency representation. An FFT size of 512 points weighed by means of a Hanning window with 50% overlapping provides adequate resolution in both, time (5.33 ms) and frequency (93.75 Hz) to allow detection and characterization of the whistle signals embedded in a noise background. The spectrogram square magnitude is then calculated and each individual time-frequency cell is normalized. The noise value estimation is performed using rank-order statistics on sliding windows in frequency.

Next steps in the processing chain correspond to image processing and start by converting the denoised matrix in an intensity image in the range 0 to 255 on which a region-based image segmentation process is applied. It is a region growing method that extracts homogeneous regions from a set of starting pixels generically referred to as 'seeds'. A set of seeds is selected for the complete image, based on the condition of exceeding a fixed threshold. From these starting points, regions corresponding to candidate whistles, are growing while surrounding pixels with a value above the noise level reference are found. This reference noise level is calculated from the complete image data.

Once the segmentation process is over, a set of candidate whistles is obtained. Normally these extracted objects span along more than one pixel in frequency for each time frame pixel and we look for obtaining frequency-time pairs that permits us to dispose of a unique bi-dimensional curve for each candidate whistle. To do so, the frequency contour is obtained based on selecting the pixel in frequency with the highest level of energy. The segmented objects can include either an individual candidate whistle or several of them mixed and therefore, in this latter case, a crossing resolution process is applied to separate the candidate whistles.

After the frequency contour extraction process is over, each candidate whistle is defined by only one frequency pixel for each time frame. Then, only candidate whistles having specific time and frequency ranges are selected and considered as whistles. These selected whistles are later characterized by means of a fixed number of RBF coefficients corresponding to the weights associated to each RBF function used to approximate the extracted curve. These coefficients can be directly used as inputs to a neural net which can perform the last stage of the pattern recognition system: the classification based on predefined criteria.

4 METHOD APPLICATION AND RESULTS

The performances of the proposed method have been evaluated both, by using simulated and real signals.

4.1 APPLICATION TO SINTHETICALLY GENERATED SIGNALS

The characteristics of the proposed method are initially shown through the processing of simulated like-whistles signals embedded in a background of Gaussian noise. A four-second length waveform is built containing two crossing quadratic swept-frequency (chirp) signals in the high range of frequencies, a linear chirp crossing with a quadratic chirp in the low range of frequencies and a logarithmic chirp alone, in the middle range of frequencies. Also three like-spurious small duration signals are incorporated to the waveform. The spectrogram from this waveform is shown in Figure 5.

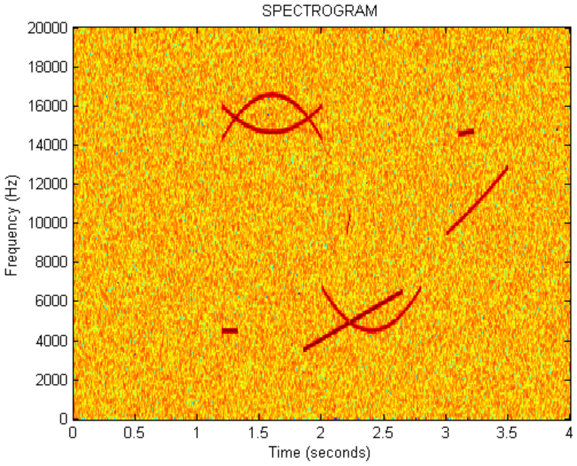

Figure 5. Spectrogram from simulated signals

The following steps of the processing chain are normalization process, to reduce the presence of noise, and the region growing based segmentation. The output of this latter step is presented in the upper left display in Figure 6. It has to be noted that crossing objects are segmented as individual ones. The next step corresponds to the contour extraction process, where each individual object is extracted and individually mapped to an image with a unique frequency value for each time frame value. After that, the selection step is performed, where only the like-whistle objects that meet the established requirements for time and frequency range are selected. Small like-artifact objects are not further taken into consideration. The output of this stage is presented in the upper right display in Figure 6. It has to be noted that although the processing module works with each extracted object individually for simplicity all the objects are shown in a unique figure. Finally, the bottom displays in Figure 6 show the extracted individual objects corresponding to the approximation functions build from the RBF weighing coefficients (RBF neural network output). These weighing coefficients characterize each object and can be further supplied to a classification system.

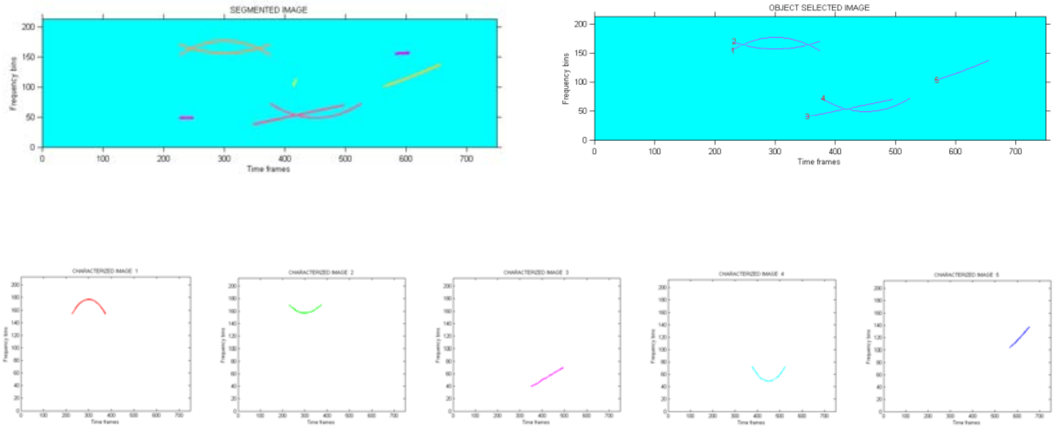

Figure 6. Outputs of the following steps: segmentation (upper-left display), objects selection (upper-right display) and approximation function built from the RBF weighing coefficients (bottom display).

4.2 APPLICATION TO REAL SIGNALS

To evaluate the performance of the proposed method, it has been applied to a set of at sea recorded signals containing whistles emitted by different kinds of dolphins and corresponding to different marine environments. The recorded whistles range from isolated ones with high signal-to-noise ratio to overlapped ones with low signal-to-noise ratio. In all cases, the method has proven to be effective to automatically detect, extract and characterize the dolphin emitted whistles. To show this, the results of the processing of two pieces of recordings are presented in Figures 7 and 8. The first of them (Figure 7) includes six dolphin whistles that are initially visualized through the spectrogram in the upper display. Displays in the middle, show the output after the segmentation stage and whistle selection stage. Finally, displays in the bottom of the image present the objects built from the RBF neural network output.

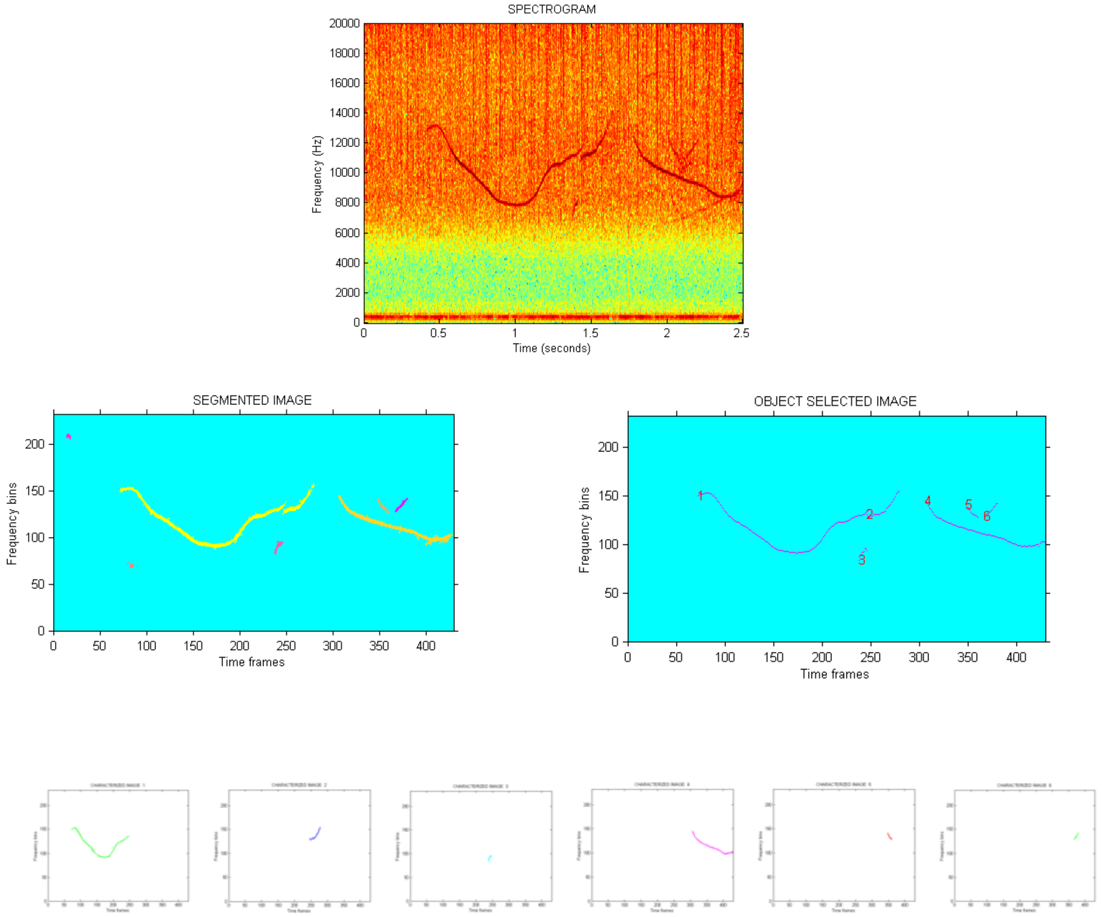

Figure 7. Results of the method applied to a first recording: spectrogram (upper display), segmentation (middle-left display), objects selection (middle-right display) and RBF neural network output (bottom display)

The second recording analyzed corresponds to a more complex marine environment, characterized by a high level of interfering noise. The same kinds of displays as for the previous case are presented in Figure 8. The only difference is that the output of the characterization stage is not presented individually, due to the high number of extracted whistles, but grouped in a sole display.

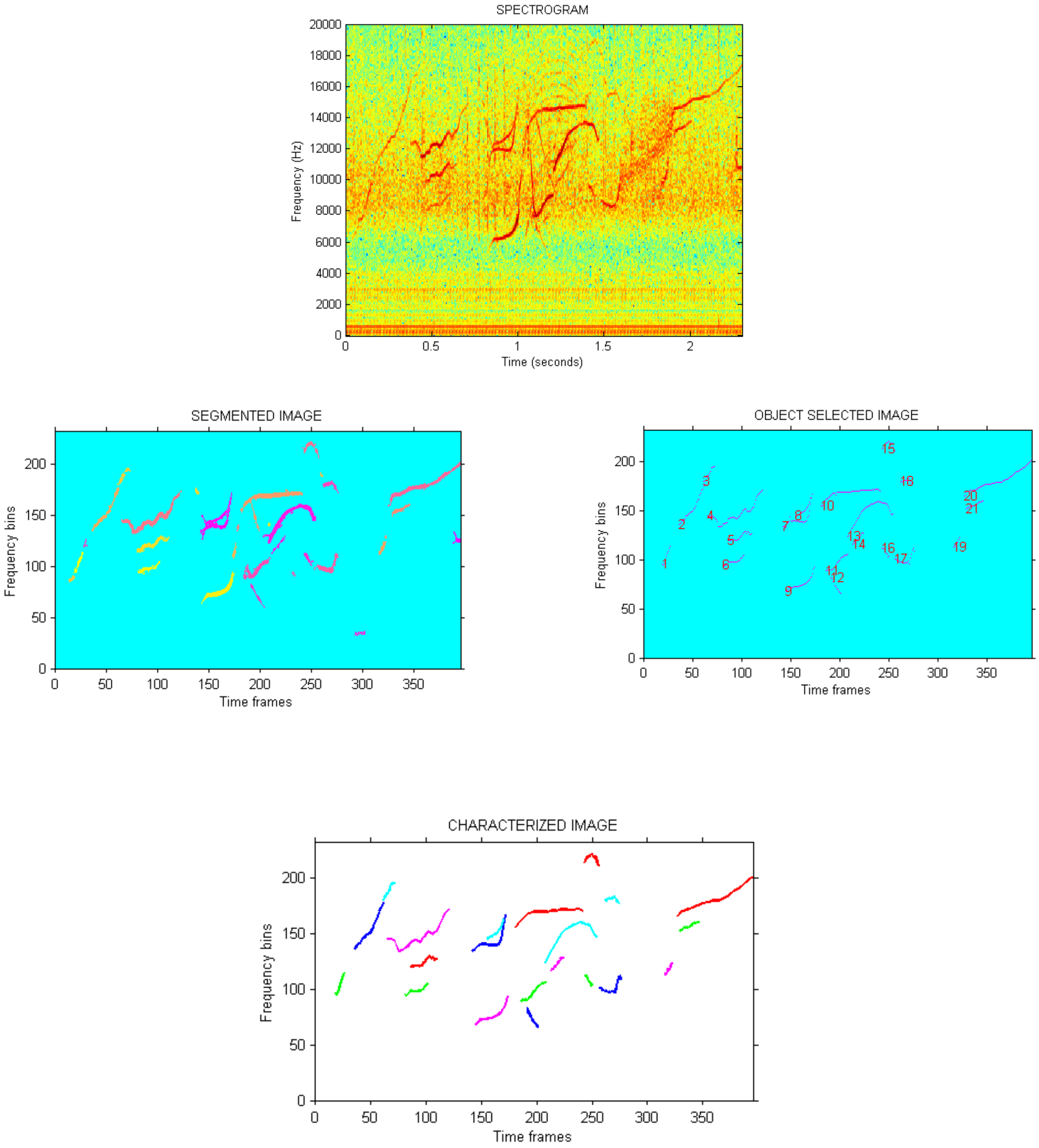

Figure 8. Results of the method applied to a second recording: spectrogram (upper display), segmentation (middle-left display), objects selection (middle-right display) and RBF neural network output (bottom display)

5 CONCLUSION AND FUTURE WORKS

An automated method for the detection and characterization of marine mammal whistles has been presented. The performances of the method have been tested on synthetic and real whistle signals from dolphins. The method combines techniques of signal processing, image processing and neural network processing and shows promising results in the line of automating the detection and identification of marine mammals. This automatic method is intended both, to reduce the sonar operator workload and to detect the presence of marine mammals in a certain area with the objective of being studied or preventing them from the exposure to potentially perturbing high levels of noise.

Present and future efforts centre on refining the algorithm in charge of solving the crossing among whistles and to link pieces of whistles split in several parts due to drops in the signal to noise ratio along the whistle. Also, it is planned to tackle the classification of the characterized signals.

REFERENCES

1. Pavan G. Acoustic risk mitigation in the Mediterranean Sea. Current situation and recommendations. In: Proceedings of the Underwater Defence Technology (UDT), Naples (Italy); June 2007.
2. Richardson WJ, Green CR, Malme CI, Thomson DH. Marine mammals and Noise. Academic Press; 1995.
3. WWL Au. The sonar of dolphins. Springer; 1993.
4. Oswald JN, Barlow J, Norris TF. Acoustic identification of nine delphinid species in the eastern tropical Pacific Ocean. Marine Mammal Sci 2003, 19:20-37.
5. González RC, Woods RE. Digital Image Processing. Prentice Hall; 2001.
6. Haykin S. Neural networks. A comprehensive foundation. Prentice Hall, 1999.